Learning Dynamics

IE 5545: Decision Analysis
Prof. Krishnamurthy Iyer*

May 2, 2026

1 Learning dynamics

We provide a brief discussion of learning dynamics — of how players might learn to play a Nash equilibrium over time, while repeatedly playing a game.

1.1 Best response dynamics

The simplest learning dynamics one might consider is where each player plays a best response to what was played by other players the last time they played the game. For example, in a Cournot competition setting where two firms compete with each other every day, each firm might decide to play the best response to the quantity the other firm produced the previous day. This process is called the best response dynamics.

Formally, we can write this as follows. Let σ_i^t denote the (mixed) strategy played by player i at time t.

1. At time $t = 0$, each player i plays an arbitrary strategy σ_i^0.
2. At any time $t \geq 1$, each player i plays a best response σ_i^t to the strategy profile of other players σ_{-i}^{t-1} at time $t-1$. In other words, we assume $\sigma_i^t \in BR_i(\sigma_{-i}^{t-1})$ for each i.

For example, in Cournot competition, the best response dynamics is as follows:

1. On day 0, firm i produces an arbitrary quantity q_i^0.
2. On days $t \geq 1$, firm i produces a quantity q_i^t that is a best response to other firms' quantity in day $t-1$, q_{-i}^{t-1}:

$$q_i^t \in BR_i(q_{-i}^{t-1}).$$

For the best response dynamics, we can show that if for each i the strategies σ_i^t converge to σ_i^* as t increases to infinity, then the limiting strategy profile $\sigma^* = (\sigma_1^*, \cdots, \sigma_N^*)$ is a (mixed) Nash equilibrium. Thus, when we have convergence, the best response dynamics provides a simple method to compute a Nash equilibrium.

For example, in the Cournot competition setting, for various cost functions, the quantities q_i^t indeed converge as $t \to \infty$. Thus, this limiting strategy profile is the (unique) Nash equilibrium of the Cournot competition.

However, one can easily come up with simple settings, where the best response dynamics does not converge, and the resulting sequence of strategies keep cycling. (Exercise: Come up with examples.)

1.2 Damped best response dynamics

Since the best response dynamics can often lead to wild jumps in the strategies being played, one can consider a "damped" version, where each player only takes a (small) step towards the best response at each time, instead of completely moving to the best response. Formally, we have the following:

1. At time $t = 0$, each player i plays an arbitrary strategy σ_i^0.
2. At any time $t \geq 1$, let $\tilde{\sigma}_i^t \in BR_i(\sigma_{-i}^{t-1})$ for each i. Each player i plays the strategy σ_i^t, where

$$\sigma_i^t = (1-\alpha)\sigma_i^{t-1} + \alpha\tilde{\sigma}_i^t,$$

for some $\alpha \in (0,1)$. Here α denotes the step size – how aggressive the player is in playing the best response. For $\alpha = 1$, we get the original best response dynamics. On the other hand, for $\alpha = 0$, each player just keeps playing the same strategy profile σ_i^0 each time period.

The damped best response dynamics has similar properties to the best response dynamics – if it converges, the limit is a Nash equilibrium. The trajectories of the damped best response dynamics are smoother than that of the best response dynamics. However, again, it is not hard to come up with examples where the damped best response dynamics does not converge. (Exercise: come up with examples.)

1.3 Fictitious play

One reason why best response dynamics fails to converge is that the players, by playing the best response to only what was played in the previous time period, forget all that was played before. In fact this "forgetting the past" feature of best response dynamics ensures that the convergence, if it happens, is fast. But this also ensures that the strategies swing wildly when there is no convergence.

One possible remedy is to consider not just what was played the previous time period, but at all the time periods before time t when choosing a best response. This leads to a learning dynamics known as *fictitious play*, introduced by George Brown [1] and further analyzed by Julia Robinson [3] in 1951. To define fictitious play, we begin by defining the empirical distribution at time t, which is given by

$$\begin{aligned}\gamma_i^t(s) &= \frac{\text{Number of times strategy } s \text{ was played by player } i \text{ before time } t}{t} \\ &= \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbf{I}\{s_i^\tau = s\}, \quad \text{for all } s \in \mathcal{S}_i.\end{aligned}$$

(Note $\mathbf{I}\{A\}$ is equal to 1 if A is true, and zero otherwise.)

The empirical distribution γ_i^t denotes a distribution over the set of pure strategies, with $\gamma_i^t(s)$ denoting the fraction of the time until time t that player i played (pure) strategy s. Let $\gamma_{-i}^t = (\gamma_1^t, \cdots, \gamma_{i-1}^t, \gamma_{i+1}^t, \cdots, \gamma_N^t)$ denote the empirical distribution of all players other than player i.

In fictitious play, each player i at each time $t \geq 1$, assumes that all other players are choosing their strategies according to the empirical distribution γ_{-i}^t, and plays a best response. Formally, for all $t \geq 1$, and for each i, we have

$$\sigma_i^t \in BR_i(\gamma_{-i}^t).$$

Again, it is easy to see that if the fictitious play converges, it must converge to a Nash equilibrium. Also, one can show that there are many cases where the best response dynamics does not converge, but under fictitious play, the empirical distribution γ^t converges to a Nash equilibrium. This in fact holds true for matching pennies. (See the python notebook on Canvas.)

Furthermore, for many classes of games, one can show that under fictitious play, the empirical distribution γ^t will converge to a mixed Nash equilibrium. One such class of games is the class of *zero-sum games*, which are all two-player games where one player's gain is another player's loss. (For details, refer [3].) Similarly, the empirical distribution under fictitious play always converges for any game (with any number of players) that is dominance solvable. (For details, refer [2].)

However, one can also show that there exist games where the fictitious play does not converge, but cycles. In fact, there is currently no known algorithm that always converges to a Nash equilibrium for a general finite game. This is an open problem and an area of active research.

1.4 Damped fictitious play

As before, we can define a damped version of fictitious play. (We leave this as an exercise.) This has similar properties to fictitious play – if it converges, the limit is a Nash equilibrium.

References

[1] George W. Brown. Iterative solution of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*. Wiley, New York, 1951.

[2] J. H. Nachbar. "Evolutionary selection" dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, 19(1):59–89, Mar 1990.

[3] Julia Robinson. An iterative method of solving a game. *Annals of mathematics*, pages 296–301, 1951.

