# Lecture 10: Learning in Games

*Lecturer: Asu Ozdaglar*

## 1 Introduction

In the next couple of lectures we will discuss *learning in games*. The reading for these classes is Fudenberg, Levine, "The Theory of Learning in Games", Chapter 1,2 (and in part, Chapter 4 and 8). The presentation of this material partially relies on the notes by Jonathan Levin.

Most economic theory relies on equilibrium analysis (such as Nash Equilibrium or some variant). A *justification* of this kind of analysis is that equilibria naturally arise as a result of learning and adaptation. This has been shown through experminatal works. In this lecture we will review some learning methods and discuss whether the learning methods lead players to Nash equilibria.

**Learning in Games** Our model of learning is as follows. We consider a strategic games that is repeatedly played. The players start with some *a-priori belief* about the strategy of the other players, and thereafter *update* this belief after each stage (i.e., after each completed game). We further assume that the players are *myopic* , i.e. the players wish to maximize their profit in each stage, based on their current belief about the strategy of the other players.

## 2 Easiest Dynamics: Best Response

We start by discussing a simple learning (i.e, update) rule which proceeds as follows: At each stage, every players chooses the best-response to the actions of all the other players in the *previous round* (i.e., the players ignore all history and simply let their belief on the strategy of the other players be based of the action of the other players in the previous round). More precisely, let $s_i^t$ denote the action of player $i$ in stage $t$. Then,

$$s_i^t = \mathsf{bestresponse}_i(s_{-i}^{t-1})$$

Clearly, it follows that if the above dynamics reaches a steady state, then this state is a Nash Equilibrium. (Recall that this was the dynamics we used in an earlier class to find the NE of the Cournot competition). Nevertheless, there are no convergence results for general games using this dynamics.

## 3 Fictitious Play

Let us now turn to a more refined learning rule which involves the players to rely on the history of games previously played to form beliefs about the opponents play. We will focus on the earliest and simplest learning rule, called the *fictitios play*.

On a high level the learning rule is as follows: Each player assumes that its opponent is using a stationnary mixed strategy. In order to estimate the mixed strategy of its opponent, each player simply observes the statistical count of each opponent strategy (and thus gets a maximum liklihood estimate the mixed strategy). A-priori beliefs are incorporated in the model by specifying "false counts" that that the players add to the historical counts.

More formally,

- Consider two players that playing the strategic game $G$ at times $t = 0, 1, 2, \ldots$. Denote the *stage payoff* by $g_i(s_i, s_{-i})$.
- Each player start with some a-priori belief on the strategy of the other player. This is specified by "false counts" (i.e., a fictitious past): $\eta_i^0(s_{-i})$.
- Let $\eta_i^t : S_{-i} \to N$ denote the number of times player $i$ has observed the action $s_{-i}$ before stage $t$, *including the fictitious past specified by the false counts.* (For example, let $n_1^0(U) = 4$, and assume that player 1 observes the sequence $\{U, D\}$. Then $n_1^2(U) = 5$.)
- Each player assumes that his opponent is using a stationnary mixed strategy $\mu_i$.
- Each player uses the following *forecast rule* to predict the strategy of the opponents at stage $t$:

$$u_i^t(s_{-i}) = \frac{\eta_i^t(s_i)}{\sum_{\bar{s}_{-i} \in S_{-i}} \eta_i^t(\bar{s}_{-i})}$$

  In other words, $i$ forecasts $(-i)$'s strategy at time $t$ to be the empirical frequence distrbution of the past play (including the fictiotious past).
- Each player chooses actions in each stage that maximixes *that stage's* expected payoff given the prediciton of the distribution of the opponents action in that stage, i.e.,

$$s_i^t \in \arg\max_{s_i \in S_i} [g_i(s_i, \mu_i^t)]$$

  (Note that it is here that we the use the condition that players are *myopic*.)

## 3.1 Example of a Fictitious Play

We will consider the fictitious play of the following game:

| | L | R |
|---|---|---|
| U | $3,3$ | $0,0$ |
| D | $4,0$ | $1,1$ |

Before we begin the analysis of the fictitious play, note that the above game only one Nash equilibrium, namely $(D, R)$, which is also the solution obtained by iterated removal of strictly dominated strategies. We proceed to the fictitious play.

**Stage 0.** At the start, prior beliefs dictate play.

| Fictitious Past | Predictions | Actions |
|---|---|---|
| $\nu_1^0 = (3, 0)$ | $\mu_1^0 = (1, 0)$ | $s_1^0 = D$ |
| $\nu_2^0 = (1, 2.5)$ | $\mu_2^0 = (1/3.5, 2.5/3.5)$ | $s_2^0 = L$ |

**Stage 1.** We proceed in the same way for Stage 1.

| Fictitious Past | Predictions | Actions |
|---|---|---|
| $\nu_1^1 = (4, 0)$ | $\mu_1^1 = (1, 0)$ | $s_1^1 = D$ |
| $\nu_2^1 = (1, 3.5)$ | $\mu_2^1 = (1/4.5, 3.5/4.5)$ | $s_2^1 = R$ |

**Stage 2.** Similarily,

| Fictitious Past | Predictions | Actions |
|---|---|---|
| $\nu_1^2 = (4, 1)$ | $\mu_1^2 = (4/5, 1/5)$ | $s_1^1 = D$ |
| $\nu_2^2 = (1, 4.5)$ | $\mu_2^2 = (1/4.5, 4.5/5.5)$ | $s_2^1 = R$ |

**Reaching a Steady State** Note that since $D$ is a dominant strategy for player 1, he always playe $D$, which means that $\mu_2^t \to (0, 1)$ with probability 1, as $t$ grows. This results in player 2 always playing $R$. We conclude that the fictitious play converges to the Nash.

**Remark** We note that a striking feature of fictitious plays is that players do not have to know anything about the opponents' payoffs. They only require a way of forming a belief about how the opponent will play.

## 3.2 The Update Rule

In this section, we discuss the update rule used in the fictitious play. As already mentionned, in each stage, each player predicts that his opponent's strategy follows the empirical frequency distribution of the past play, including the fictitious past.

This type of update rule has also been considered as the maximum likelihood estimation. That is, a player has observations of his opponent's behavior and would like to use these observations to determine his opponent's mixed strategy. He does so by picking the mixed strategy which maximizes the likelihood of the observations. Indeed, the simple manner in which our frequency counts are updated does exactly this!

Another way for a player to predict the opponent's strategy based on empirical observation is through the *Bayesian approach*. Here, instead of starting with a *fictitious past*, we assume that players have an a-priori belief, i.e. a probability distribution, on their opponents' mixed strategies. Now, based on the empirical observations, each player updates their own belief about their opponents based on Bayes' rule. More formally,

- Each player $i$ begins with a *prior* belief on $\Sigma_{-i}$. That is, each player begins with $\rho_i$ which is a probability distribution over $\Sigma_{-i}$. ( Recall that $\Sigma_{-i}$ is the set of probability distributions over the actions $S_{-i}$. )
- After round $t$, each player updates their belief about opponent strategies for round $t+1$ using Bayes rule to compute the posterior.

  Even more formally, consider a sequence of $n$ independent and identically distributed (i.i.d. for short) trials where in each period, one of $k$ outcomes occurs. Let $p_z$ denote the probability of outcome $z$.

Denote the outcome of the $n$ trials by $K$, and the number of times that the outcome $z$ occurs by $K_z$. (In other words, $K$ is a vector of the values, and $K_z$ is the number of times that outcome $z$ occurs in $K$. This is a slight abuse of standard notation.)

Define

$$f(K|p) = \left(\frac{n!}{K_1! \cdots K_k!}\right) p_1^{K_1} \cdots p_k^{K_k}$$

to be the probability that vector $K$ occurs given the probability distribution $p$. Then, using Bayes rule, compute

$$f(p|K) = \gamma f(K|p) f(p)$$

where $\gamma$ is a normalization factor equal to $\int f(K|p) f(p) dp$.

Finally, compute the mixed strategy of the opponent by taking the expectation of $p$.

Below, we show that the two approaches of using the maximum likelihood and the Bayesian approach can be reconciled if we assume that the players' a-priori beliefs are Dirichlet distributions.

**Definition 1 (Dirichlet)** *Let $p = (p_1, \ldots, p_k)$ be a vector of probabilities where $p_i$ represents the probability of event $i$. A* Dirichlet *prior $\rho$ with parameter $\alpha = (\alpha_1, \ldots, \alpha_k)$ is described by the probability density function:*

$$f(p) = c \prod_{z=1}^{k} p_z^{\alpha_z - 1}$$

*where $c$ is a normalization constant.*

The marginal distribution of $p_z$ is

$$E[p_z] = \frac{\alpha_z}{\sum_{w=1}^{k} \alpha_w}$$

It follows that

$$f(p|K) = \prod_{z=1}^{K} p_z^{(\alpha_z + K_z) - 1}$$

Finally, to compute the expectation of $p$, we use the following rule:

$$E[p_z|K] = \int_{p_z=0}^{1} p_z \int_{p_1} \int_{p_2} \cdots \int_{p_{z-1}} f(p|K) = \frac{\alpha_z + K_z}{\sum_{w=1}^{k} \alpha_w + K_w}$$

which is the same rule used in the maximum likelihood method.

## 3.3 Remark about Fictitious Play

During fictitious play, player $i$ assumes (incorrectly) that player $j$ will play a stationary mixed strategy against him. Since player $j$ is *also* applying a learning rule, this implies that player $i$'s belief about $j$ will be wrong, even though $i$ updates correctly from his prior. Nonetheless, fictitious play will sometimes converge to an equilibrium.

# 4 Convergence of Fictitious Play

In this section, we establish a proposition regarding the convergence of Fictitious Play. First, we define the notion of convergence for a sequence of strategies.

**Definition 2** *A sequence of strategies $\{s^t\}$ converges to $s$ if there exists a $T$ such that for all $t > T$ $s^t = s$.*

**Proposition 1** *Let $\{s^t\}$ be the sequence of strategies defined by a Fictitious Play. Then*

*(a) if $\{s_t\}$ converges to $s$, $s$ is a Nash equilibrium, and*

*(b) if $s^T = s^*$ where $s^*$ is a strict Nash equilibrium, $s^\ell = s^*$ for all $\ell > T$.*

*Proof:* For (a), if fictitious play remains at a pure strategy profile $s$, then eventually, belief forecasts about the opponents strategy will also converge to $s$. Given that player $i$ believes all other players play $s_{-i}$, and given that fictitious play dictates that player $i$ chooses $s_i$ to maximize his utility given his belief, it follows that $s$ is a Nash equilibrium.

For (b), if $s^t = s^*$, we show that $s^{t+1} = s^*$. Recall that the update rule is such that

$$\mu_i^{t+1} = (1-\alpha)\mu_i^t + \alpha s_{-i}^t = (1-\alpha)\mu_i^t + \alpha s_{-i}^*$$

where $\alpha$ is a normalization factor equal to $\alpha = \frac{1}{\left(\sum_{s_{-i}} \nu_i^t(s_{-i})\right)+1}$.

Consider the utility function,$g_i$, for stage $t+1$. By the linearity of expectations of $\mu$, we have

$$g_i(a_i, \mu_i^{t+1}) = (1-\alpha)g_i(a_i, \mu_i^t) + \alpha g_i(a_i, s_{-i}^*)$$

Since $s^*$ maximizes both terms (since it is a strict Nash), then $s_i^*$ will also be played at stage $t+1$ which completes the proof. $\square$