In this lecture, we develop the notion of "adaptive learning" as proposed by Milgrom and Roberts [1]. Although the learning definition they give is of interest in its own right, it primarily derives power in the case of dominance solvable games, or for games where there is a straightforward characterization of the set of strategies surviving iterated strict dominance (hereafter ISD).

Throughout the lecture we consider a finite N-player game, where each player i has a finite pure action set A_i; let $A = \prod_i A_i$. We let a_i denote a pure action for player i, and let $s_i \in \Delta(A_i)$ denote a mixed action for player i. We will typically view s_i as a vector in $\mathbb{R}^{A_i}$, with $s_i(a_i)$ equal to the probability that player i places on a_i. We let $\Pi_i(\boldsymbol{a})$ denote the payoff to player i when the composite pure action vector is $\boldsymbol{a}$, and by an abuse of notation also let $\Pi_i(\boldsymbol{s})$ denote the expected payoff to player i when the composite mixed action vector is $\boldsymbol{s}$. We let $BR_i(\boldsymbol{s}_{-i})$ denote the best response mapping of player i; here $\boldsymbol{s}_{-i}$ is the composite mixed action vector of players other than i.

We will need some additional notation involving ISD. Given $T \subset \prod_i A_i$, we define $U_i(T)$ as follows:

$$U_i(T) = \{a_i \in A_i : \text{for all } s_i \in \Delta(A_i), \text{ there exists } \boldsymbol{a}_{-i} \in T_{-i} \text{ s.t. } \Pi_i(a_i, \boldsymbol{a}_{-i}) \geq \Pi_i(s_i, \boldsymbol{a}_{-i})\}.$$

Here $T_{-i} = \prod_{j \neq i} T_j$, where T_j is the projection of T onto A_j. In other words, $U_i(T)$ is the set of pure strategies of player i that are *not dominated by any mixed strategy*, given that all other players play using action vectors in T_{-i}. We let $U(T) = \prod_i U_i(T)$. We also use $U^k(T)$ to denote the set of pure strategies remaining after k applications of U to the set T, with U^0 equal to the identity map.

It is straightforward to check the following claims (see Lemmas 1 and 2 of [1]):

1. *Monotonicity*: If $T \subset T'$, then $U(T) \subset U(T')$.
2. *Decreasing sequence property*: $U^{k+1}(A) \subset U^k(A)$ for all k. (Note that this need not be true if we iterate U starting from a set *strictly smaller* than the entire strategy space A, since for an arbitrary set T we need not have $U(T) \subset T$.)

In light of the second claim, we let $U^\infty(A) = \bigcap_{k \geq 0} U^k(A)$. Note that this is the set of *strategies surviving ISD*.

1 Adaptive Learning

Milgrom and Roberts define their notion of learning in terms of an arbitrary (discrete-time) sequence of action vectors $\{\boldsymbol{a}^t\}$. The idea is that if player i is adapting to his opponents' play, then a_i^t should "eventually" be an undominated strategy, if player i assumes his opponents will only play actions they have played in the "recent" past.

Formally, they say the sequence $\{a_i^t\}$ is *consistent with adaptive learning for player* i if for all $t' \geq 0$, there exists $\bar{t} \geq t'$ such that for all $t \geq \bar{t}$, there holds $a_i^t \in U(\{\boldsymbol{a}^s : t' \leq s < t\})$. The

sequence $\{\boldsymbol{a}^t\}$ is *consistent with adaptive learning* if $\{a_i^t\}$ is consistent with adaptive learning for all players i.

In the definition, the value t' defines a "recent past", and the value $\bar{t}$ defines a reasonable "adaptation period". This is significantly more general than fictitious play: there is no requirement that every play of player i should be undominated given the entire past history. Rather, after looking at the play of his opponents, player i should eventually play strategies that are undominated, when ignoring strategies of players $j \neq i$ that have not been played for a sufficiently long time.

As Milgrom and Roberts note, schemes like fictitious play, best response dynamics, and even Bayesian learning are all consistent with adaptive learning. To get a feel for this, we consider a class of learning algorithms where player i plays a best response to *some* probability distribution over the past history of his opponents' play. Formally let $h^t = (\boldsymbol{a}^0, \ldots, \boldsymbol{a}^{t-1})$ be the *history* up to time t, and let $\mu_i(h^t)$ denote the *belief* of player i; this is a probability distribution over $\prod_{j \neq i} A_j$, and forecasts the action vector player i expects his opponents to play at time t. Note that μ_i can be derived in many ways from past history: it may the product of the empirical distributions of $\{a_j^t\}$ for players $j \neq i$ (as in fictitious play); it may be the empirical *joint* distribution of $\{\boldsymbol{a}_{-i}^t\}$; it may be an exponentially weighted moving average of $\{\boldsymbol{a}_{-i}^t\}$; it may place unit mass on the *last* play of i's opponents, $\boldsymbol{a}_{-i}^{t-1}$ (as in the standard best response dynamics); etc.

Since $\mu_i(h^t)$ is a probability distribution over opponents' actions, we can view $\mu_i(h^t)$ as the *predicted mixed strategy* of player i's opponents. We consider dynamics where each player i chooses a_i^t as a *best response* to $\mu_i(h^t)$. Formally, a *best response dynamic (BRD) with forecasters* $\boldsymbol{\mu}$ is a sequence of action vectors $\boldsymbol{a}^t$ such that for all periods $t > 0$ and all players i, there holds $a_i^t \in BR_i(\mu_i(h^t))$.

Of course, the formulation so far is sufficiently general that the belief function μ_i could even be trivial, and not respond at all to opponents' past play. It is clear that such a belief function could not in general give rise to a BRD that is consistent with adaptive learning: player i may be playing strategies that are dominated given his opponents' past play. To counteract this possibility, we say that the forecaster μ_i is *adaptive* if, for any action a_j of player $j \neq i$ that is only played finitely often in the sequence $\{\boldsymbol{a}^t\}$, the belief $\mu_i(h^t)(a_j)$ converges to zero; that is, if player j only plays a_j finitely many times, then eventually player i's belief must place zero weight on player j playing a_j^t. (Note that we are only considering here the *marginal* belief over player j's actions; e.g., in fictitious play, it is possible that individual players $j \neq i$ play a_j infinitely often, even though the composite action vector $\boldsymbol{a}_{-i}$ is never played.)

Milgrom and Roberts present the following theorem (Theorem 8 in [1]); note that it makes use of the finiteness of the action spaces.

Theorem 1 *Suppose that $\{\boldsymbol{a}^t\}$ is a BRD with forecasters $\boldsymbol{\mu}$, and each forecaster μ_i is adaptive. Then $\{\boldsymbol{a}^t\}$ is consistent with adaptive learning.*

Proof. Suppose that the theorem is false. Let $T_j \subset A_j$ be the set of all pure actions played infinitely often by player j; and let $T_{-i} = \prod_{j \neq i} T_j$. If the theorem is false, there must exist a player i, and a sequence of times t_k and mixed strategies s_i^k such that:

$$\Pi(s_i^k, \boldsymbol{a}_{-i}) > \Pi(a_i^{t_k}, \boldsymbol{a}_{-i}),$$

for all action vectors $\boldsymbol{a}_{-i} \in T_{-i}$. In other words, a_i^t must be strictly dominated infinitely often, assuming opponents play actions drawn from T_{-i}. Without loss of generality (taking subsequences if necessary), we can assume that $a_i^{t_k} = a_i$ for all k (since player i has finitely many actions), and that $\mu_i(h^{t_k}) \to \mu_i^*$ as $k \to \infty$. Under the first assumption we can also assume without loss of generality that $s_i^k = s_i$ for all k.

Since there are only finitely many action vectors, there exists $\varepsilon > 0$ such that:

$$\Pi(s_i^k, \boldsymbol{a}_{-i}) > \Pi(a_i^{t_k}, \boldsymbol{a}_{-i}) + \varepsilon,$$

for all action vectors $\boldsymbol{a}_{-i} \in T_{-i}$. Note that μ_i^* has support only in T_{-i}, by the assumption that the forecaster μ_i is adaptive. Therefore we have:

$$\Pi(s_i, \mu_i^*) > \Pi(a_i, \mu_i^*) + \varepsilon.$$

But then for all sufficiently large k we have:

$$\Pi(s_i^k, \mu_i(h^{t_k})) > \Pi(a_i^{t_k}, \mu_i(h^{t_k})),$$

which contradicts the assumption that $a_i^{t_k}$ was a best response at time t_k. This establishes the theorem. □

The preceding theorem shows that "consistent with adaptive learning" is a broad enough concept to capture any of the basic learning models we have studied so far.

2 Convergence to U^∞

As we might expect, if play is consistent with adaptive learning, then players eventually play only actions that survive ISD. We start with the following result (Theorem 5 in [1]).

Proposition 2 *Suppose $\{\boldsymbol{a}^t\}$ is consistent with adaptive learning. Then for all k there exists a time t_k such that $\boldsymbol{a}^t \in U^k(A)$ for all $t \geq t_k$.*

Proof. The conclusion is trivially true for $k = 0$; assume it holds for $k = n$. We show it holds for $k = n + 1$.

Let $\bar{t}_n$ be the threshold in the definition of consistency with adaptive learning, when $t' = t_n$; that is, choose $\bar{t}_n \geq t_n$ such that for all $t \geq \bar{t}_n$, there holds:

$$a_i^t \in U(\{\boldsymbol{a}^s : t_n \leq s < t\}),$$

for all players i. Now observe that $\{\boldsymbol{a}^s : t_n \leq s < t\} \subset U^n(A)$, by the inductive hypothesis. Thus:

$$U(\{\boldsymbol{a}^s : t_n \leq s < t\}) \subset U(U^n(A)) = U^{n+1}(A),$$

so the result holds if we choose $t_{n+1} = \bar{t}$. Q.E.D. □

A simple consequence of the preceding proposition is that play eventually converges to the set of actions that survive ISD.

Theorem 3 *Suppose $\{\boldsymbol{a}^t\}$ is consistent with adaptive learning. Let $A_i^\infty \subset A_i$ be the set of actions played infinitely often by player i, and let $A^\infty = \prod_i A_i^\infty$. Then $A^\infty \subset U^\infty(A)$.*

In games with finite action spaces, the theorem implies that there exists a finite time t after which players only play actions that survive ISD. As a consequence, we have the following corollary.

Corollary 4 *Suppose that $U^\infty(A) = \{\boldsymbol{a}^*\}$; i.e., the game is dominance solvable. Then $\boldsymbol{a}^t$ converges to $\boldsymbol{a}^*$ if and only if $\{\boldsymbol{a}^t\}$ is consistent with adaptive learning.*

Here "convergence" means that there exists t^* such that for all $t \geq t^*$, we have $\boldsymbol{a}^t = \boldsymbol{a}^*$. The proof is immediate: from Theorem 3, it is clear that if $\{\boldsymbol{a}^t\}$ is consistent with adaptive learning, then it converges to $\boldsymbol{a}^*$. Conversely, if play is stationary at $\boldsymbol{a}^*$ after some time, then play is trivially consistent with adaptive learning. (Milgrom and Roberts prove a slightly more sophisticated version of this theorem that holds when action spaces are not finite; see Theorem 7 in [1].)

In particular, note that the preceding corollary together with Theorem 1 establishes convergence of fictitious play in games that are dominance solvable. This also explains the close relationship between adaptive learning and *supermodular games*. For supermodular games, we know the set of actions surviving ISD is upper and lower bounded by the "largest" and "smallest" pure NE. If there is a unique NE in a supermodular game, then it is dominance solvable, so any dynamic consistent with adaptive learning converges. Further, for a general supermodular game, play eventually lies between the largest and smallest pure NE.

References

[1] P. Milgrom and J. Roberts. Adaptive and sophisticated learning in normal form games. *Games and Economic Behavior*, 3:82–100, 1991.