CSCE 475/875 Multiagent Systems

Handout 10: Rational Learning (Bayesian Learning)

September 14, 2017

(Based on Shoham and Leyton-Brown 2011)

Introduction

Rational learning (also sometimes called *Bayesian learning*) adopts the same general model-based scheme as fictitious play. ***Unlike fictitious play, however, it allows players to have a much richer set of beliefs about opponents' strategies***.

As in fictitious play, each player begins the game with some prior beliefs. After each round, the player uses *Bayesian updating* to update these beliefs.

Let S_{-i}^{i} be the set of the opponent's strategies considered possible by player i, and H be the set of possible histories of the game. Then we can use Bayes' rule to express the probability assigned by player i to the event in which the opponent is playing a particular strategy $s_{-i} \in S_{-i}^{i}$ given the observation of history $h \in H$, as

$$P_i(s_{-i}|h) = \frac{P_i(h|s_{-i})P_i(s_{-i})}{\sum_{s'_{-i} \in S_{-i}^{i}} P_i(h|s'_{-i})P_i(s'_{-i})}$$

Definition 7.3.1 (Absolute continuity) Let X be a set and let $\mu, \mu' \in \prod(X)$ be probability distributions over X. Then the distribution μ is said to be absolutely continuous with respect to the distribution μ' iff for $x \subset X$ that is measurable it is the case that if $\mu(x) > 0$ then $\mu'(x) > 0$.

Note: Rational learning is a very intuitive model of learning, but its analysis is quite involved. Roughly speaking, players satisfying the assumptions of the rational learning model will have beliefs about the play of the other players that converge to the truth, and furthermore, players will in finite time converge to play that is arbitrarily close to the Nash equilibrium.

***Note**:* Thus a player's beliefs will eventually converge to the truth if it is using Bayesian updating, is playing a best response strategy, and the play predicted by the other players' real strategies is absolutely continuous with respect to that predicted by its beliefs. In other words, it will correctly predict the *on-path portions* of the other players' strategies. This says that if an agent keeps playing its best response based on its beliefs of other agents, then eventually it will start to play the best way that it can be.

Note: This result does *not* state that players will learn the true strategy being played by their opponents. However, a player's beliefs must converge to the truth even when its strategy space is incorrect (does not include the opponent's actual strategy), as long as they satisfy the absolute continuity assumption.

Note: In other words, if *utility-maximizing players start with individual subjective beliefs with respect to which the true strategies are absolutely continuous*, then in the long run, their behavior must be essentially the same as a behavior described by an ε-Nash equilibrium.